Text Summarization

• **Goal**: produce an abridged version of a text that contains information that is important or relevant to a user.

• **Summarization Applications**
  • _outlines or abstracts_ of any document, article, etc
  • _summaries_ of email threads
  • _action items_ from a meeting
  • _simplifying_ text by compressing sentences

What to summarize?  
Single vs. multiple documents

• **Single-document summarization**
  • Given a single document, produce
    • abstract
    • outline
    • headline

• **Multiple-document summarization**
  • Given a group of documents, produce a gist of the content:
    • a series of news stories on the same event
    • a set of web pages about some topic or question
Query-focused Summarization & Generic Summarization

• **Generic summarization:**
  • Summarize the content of a document

• **Query-focused summarization:**
  • Summarize a document with respect to an information need expressed in a user query.
  • A kind of complex question answering:
    • Answer a question by summarizing a document that has the information to construct the answer

Summarization for Question Answering: Snippets

• **Create snippets** summarizing a web page for a query
  • Google: 156 characters (about 26 words) plus title and link
Summarization for Question Answering: Multiple documents

Create answers to complex questions summarizing multiple documents.

- Instead of giving a snippet for each document
- Create a cohesive answer that combines information from each document

Extractive summarization & Abstractive summarization

- Extractive summarization:
  - create the summary from phrases or sentences in the source document(s)
- Abstractive summarization:
  - express the ideas in the source documents using (at least in part) different words
Simple baseline: take the first sentence

ROUGE (Recall Oriented Understudy for Gisting Evaluation)  
Lin and Hovy 2003

• Intrinsic metric for automatically evaluating summaries
  • Based on BLEU (a metric used for machine translation)
  • Not as good as human evaluation
  • But much more convenient (and some correlation with humans)

• Given a document \( D \), and an automatic summary \( X \):
  1. Have \( N \) humans produce a set of reference summaries of \( D \)
  2. Run system, giving automatic summary \( X \)
  3. What percentage of the bigrams from the reference summaries appear in \( X \)?

\[
ROUGE = 2 - \frac{\sum_{s \in \{\text{RefSummaries}\}} \sum_{\text{bigrams} i \in S} \min(\text{count}(i,X),\text{count}(i,S))}{\sum_{s \in \{\text{RefSummaries}\}} \sum_{\text{bigrams} i \in S} \text{count}(i,S)}
\]
A ROUGE example:
Q: “What is water spinach?”

Human 1: Water spinach is a green leafy vegetable grown in the tropics.
Human 2: Water spinach is a semi-aquatic tropical plant grown as a vegetable.
Human 3: Water spinach is a commonly eaten leaf vegetable of Asia.

• System answer: Water spinach is a leaf vegetable commonly eaten in tropical areas of Asia.

• ROUGE-2 = \[ \frac{3 + 3 + 6}{10 + 9 + 9} = \frac{12}{28} = .43 \]